ISSN 1561-5359. Штучний інтелект, 2017, № 1

УДК 004.89, 004.93

А.В. Агарков
Інститут проблем штучного інтелекту МОН України і НАН України, Україна
пр. Академіка Глушкова, 40, м. Київ, 03680

ЗАСТОСУВАННЯ СТРУКТУРНОГО ОПИСУ ЗОБРАЖЕНЬ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ВІДЕОПОСЛІДОВНОСТЕЙ

A.V. Agarkov
Institute of Artificial Intelligence MES of Ukraine and NAS of Ukraine, Ukraine
40, Academician Hlushkov av., Kyiv, 03680

APPLICATION OF IMAGES STRUCTURAL DESCRIPTION FOR SOLVING PROBLEMS OF INTELLECTUAL ANALYSIS OF VIDEO SEQUENCES

In this paper, we consider the use of the description of images and video sequences in the form of a set of structural elements for solving problems of detection, tracking and recognition of moving objects. Theoretical studies of this problem have been carried out. A formal description of images and video sequences in the form of a number of structural elements is given, a definition of the description of detected and tracked objects is given, properties of structural elements belonging to one object, properties of detected objects description, properties of tracked objects descriptions, properties of the transformation/modification function of tracked objects description from frame to frame, which is necessary for tracking, are considered.

Key words: object tracking, detection of moving objects, structural elements, structural description of images.

У даній роботі розглянуто використання опису зображень і видеопослідовностей у вигляді безлічі структурних елементів для вирішення завдань детектування, відстеження та розпізнавання рухомих об'єктів. Проведено теоретичні дослідження даної проблеми – даний формальний опис зображень і видеопослідовностей у вигляді безлічі структурних елементів, дано визначення опису виділених і об'єктів, що відслідковуються, розглянуто властивості структурних елементів, що належать одному об'єкту, властивості опису виділених об'єктів, властивості описів об'єктів, що відслідковуються, властивості функції перетворення/модифікації опису об'єктів, що відслідковуються від кадра до кадра, яка необхідна для здійснення трекінгу.

Ключові слова: трекінг об'єктів, детектування рухомих об'єктів, структурні елементи, структурний опис зображень.

Вступ

Інтелектуальний аналіз відеоданих передбачає використання методів комп'ютерного зору для автоматичного виявлення, відстеження і розпізнавання об'єктів у полі зору відеокамери. Дані методи і технології дозволяють аналізувати відео без участі людини. Потреба в системах, здатних в автоматичному режимі здійснювати інтелектуальний аналіз відеоданих, зараз дуже велика внаслідок великої кількості відеоматеріалу, який залишається без аналізу.

Одним з найбільш ранніх способів виділення об'єктів на зображеннях кадрів відеопослідовності є метод віднімання фону [1-5]. Даний метод заснований на припущенні, що яскравість окремого пікселя, що належить фону, залишається приблизно постійною. Її зміни обумовлені нестабільністю світлочутливої матриці відеокамери і описуються нормальним законом розподілу. Якщо яскравість пікселя відхиляється від допустимого діапазону, то це значить, що даний піксель належить вже не фону, а об'єкту переднього плану. Виділені таким способом пікселі об'єднуються у зв'язані області, які відповідають об'єктам переднього плану. Для

кожної області формується дескриптор, який використовується для порівняння з областями на попередньому кадрі, що дозволяє провести ідентифікацію об'єктів на поточному кадрі. Даний метод є дуже простим у реалізації, але досить ефективним, щоб бути використаним для вирішення різних задач комп'ютерного зору [6-9]. Але його застосування вимагає достатньо чіткої віддільності об'єктів від фону і мінімальної кількості перекриттів між об'єктами.

У роботі [10] для трекінгу об'єктів використовується аналіз міжкадрової зміни в області, в якій знаходиться об'єкт, і зіставлення областей для локалізації об'єкта на наступному кадрі. У роботах [11-13] для визначення положення об'єкта на наступному кадрі використовується кореляційний підхід. Для цього формується адаптивний шаблон, який підлаштовується під зміни об'єкту. Але при постійних змінах об'єктів спостерігається дрейф – сповзання області відслідковування з об'єкту.

У роботі [14] для локалізації об'єкта використовується узагальнене перетворення Хафа. Даний метод стійкий до зміни об'єкта, але вимагає хорошої віддільності об'єкта від фону.

У роботах [15,16] для виділення і відстеження об'єктів використовується сегментація оптичного потоку. Для точок кожного кадру відеопослідовності визначається оптичний потік – міжкадрове зміщення для кожної розглянутої точки. Таким чином формуються траєкторії точок, які переміщуються від кадра до кадра. Окремим об'єктам відповідають множини точок з паралельною траєкторією. Виділення зв'язкових підмножин траєкторій дозволяє виділяти об'єкт відразу на послідовності кадрів. Якісний розрахунок оптичного потоку вимагає достатньо великого обсягу ресурсів, що не дозволяє використовувати метод для задач реального часу, проте підхід, заснований на сегментації множини траєкторій, заслуговує на увагу.

Роботи [17-20] присвячені розвитку методу FoT (Flock of Trackers – зграя трекерів). Суть даного методу полягає в тому, що окремий об'єкт відстежує від кадра до кадра множину трекерів. Результуюче міжкадрове зміщення для об'єкта, що відслідковується, формується на основі аналізу даних, отриманих від даної множини. Різниця між реалізаціями даного методу полягає у способі розрахунку результуючого значення міжкадрового зміщення для об'єкта. Перевагою даного методу є висока швидкість роботи, що дозволяє використовувати його в задачах реального часу. Однак даний метод не позбавлений типових недоліків багатьох трекерів – дрейф при поворотах об'єкта навколо осі перпендикулярної оптичній осі камери.

У роботах [21-31] використовується трекінг за допомогою детектування. У режимі реального часу здійснюється навчання детектора для цільового об'єкта. При зміні об'єкта детектор донавчається, що дозволяє утримувати об'єкт тривалий час при різних його змінах, знаходити його після втрати (зникнення об'єкта в кадрі).

Роботи [21-25] присвячені розвитку методу TLD (Tracking-Learning-Detection). Метод TLD заснований на використанні для пошуку положення об'єкта на наступному кадрі двох механізмів – пошук міжкадрового зміщення для об'єкта на основі оптичного потоку і пошук положення об'єкта на основі використання детектора. Міжкадрове зміщення об'єкта визначається на основі аналізу оптичного потоку множини точок, що належать об'єкту. Детектор заснований на використанні Хара-подібних властивостей і методу навчання AdaBoost. При ініціалізації трекера формується навчальна вибірка, що складається із зображень об'єкта і фону, які витягуються з поточного кадру. На основі даної вибірки проводиться навчання

класифікатора методом AdaBoost. Оскільки навчальна вибірка не велика, то навчання не потребує значних обчислювальних ресурсів і проводиться в режимі реального часу. Якщо результати пошуку міжкадрового зміщення і детектора відрізняються – навчальна вибірка поповнюється новими зображеннями об'єкта і фону і проводиться перенавчання детектора.

У роботах [26,27] для детектування використовується більш складна, ніж у TLD схема навчання детектора, яка використовує кілька зсунутих відносно один одного зображень об'єкта на кадрі, що дозволяє зменшити ймовірність дрейфу.

Метод, описаний у роботах [28,29], використовує для детектування об'єкта розкладання за власними векторами зображення об'єкта, що дозволяє відстежити і врахувати в просторі ознак основні зміни зовнішнього вигляду об'єкта.

Роботи [30,31] присвячені адаптації методу TLD для трекінгу декількох об'єктів, що перебувають одночасно в кадрі. Для прискорення роботи і підвищення якості введена обмежена область пошуку детектором об'єкта на поточному кадрі. Для підвищення точності визначення положення об’єкта використовуються два методи пошуку міжкадрового зміщення об'єкта і додатковий попередньо навчений детектор, призначений для всіх об'єктів даного класу.

Дані методи потребують зовнішньої ініціалізації об'єкта з мінімальним захопленням фону. У даних методах не передбачені зміни розмірів об'єкта, що може трапиться при зміні відстані від об'єкта до камери. При збільшенні розміру об'єкта знижується якість детектування об'єкта – об'єкт розбивається на кілька областей детекції. Починає грати велику роль відносний рух частин об'єкта щодо один одного, що знижує якість трекінгу при використанні методу адаптивного шаблону. У такій ситуації є розумним відстежувати переміщення окремих частин незалежно одна від одної. Також необхідно об'єднувати відслідковані частини одного об'єкта в одне ціле, щоб стежити за самим об'єктом, а не його окремими частинами.

Представлення зображення і відеопослідовності у вигляді множини структурних елементів дозволяють повною мірою реалізувати даний підхід для вирішення завдання відстеження великих об'єктів. Кожен структурний елемент відповідає певній області зображення, має координати в площині зображення і дескриптор, що описує характеристики яскравості/текстурні властивості даної області. Просторові і фазові відносини між структурними елементами визначають структуру зображення і зображень окремих об'єктів.

Дана робота є узагальненням попередніх дослідів [32-34], у яких застосовується представлення зображення як множини структурних елементів. Запропонований метод дозволяє здійснювати виділення й відстеження об'єктів, що рухаються, за допомогою єдиного процесу.

Структурний опис зображень і відеопослідовностей

Як правило, зображення розглядається як двовимірна дискретна обмежена функція. Для вирішення задачі виділення і відстеження рухомих об'єктів пропонується замість даної функції використовувати представлення зображення у вигляді множини структурних елементів. Кожен структурний елемент відповідає певній області зображення, має координати в площині зображення і дескриптор, що описує характеристики яскравості і текстурні властивості даної області. Просторові і фазові відносини між структурними елементами визначають структуру зображення і зображень окремих об'єктів. При використанні даного подання зображень окремим об'єктам відповідають підмножини структурних елементів. Таким чином, завдання

виділення цільових об'єктів зводиться до пошуку відповідних підмножин структурних елементів.

Для опису зображення в кожній точці введемо функцію яскравості $I(x,y)$, де *x*, *y* координати пікселів зображення.

Перехід від функції яскравості до множини умовно позначимо як

$$I(x,y) \Rightarrow E(I) = \{e_i = (x_i, y_i, \vec{s}_i, D_i)\}_i,$$

де $I(x,y)$ – початкове зображення, представлене як функція яскравості I в кожній точці (x,y), що належить зображенню; $e_i = (x_i, y_i, \vec{s}_i, D_i)$ – *i*-тий структурний елемент, який має координати центру в площині зображення (x_i, y_i), D_i – текстурний дескриптор, що описує характеристики яскравості і/або текстурні властивості околиці точки (x_i, y_i), $\vec{s}_i$ – зміщення між кадрами *i*-того структурного елемента (вектор, який визначає різницю в координатах центру структурного елемента на двох послідовних зображеннях відеоряду). Параметр $\vec{s}_i$ визначений тільки для структурних елементів зображень, що належать відеоряду.

Оскільки кожен структурний елемент має певні координати на зображенні, то всі вони мають властивість локальності, тобто мають певне місце розташування.

Відеоряд $V(x,y)$ представляється, зазвичай, як послідовність зображень

$$V(x,y) = \{I_l(x,y), l = 1,..N_V\},$$

де $I_l(x,y)$ – *l*-те зображення (кадр) відеоряду, N_V – кількість кадрів у відеоряді.

Оскільки для відеоряду один і той же структурний елемент належить цілій множині зображень, то перехід до подання зображення у вигляді множини структурних елементів слід позначити:

$$V(x,y) \Rightarrow E(V) = \cup E(I_l) = \{e_i = (\{(x_i, y_i)^l\}_l, \{\vec{s}_i^{\,l}\}_l, \{D_i^l\}_l)\}_i,$$

де $\{(x_i, y_i)^l\}_l$ – множина координат центру *i*-того структурного елементу на кадрах відеопослідовності, $\{D_i^l\}$ – множина текстур дескрипторів для центру *i*-того структурного елемента на кадрах відеопослідовності, $\{\vec{s}_i^{\,l}\}_l$ – множина міжкадрових зміщень для *i*-того структурного елемента на кадрах відеопослідовності. З визначення зміщення між кадрами відношення між ним і координатами центру структурного елемента на двох послідовних кадрах:

$$\vec{s}_i^{\,l} = (x_i, y_i)^l - (x_i, y_i)^{l-1}. \tag{1}$$

Оскільки передбачається, що окремий структурний елемент відповідає певній області об'єкта або фону на зображенні, то для текстурних дескрипторів структурних елементів справедливе відношення:

$$\forall l,k\, D_i^l \approx D_i^k : l \neq k, e_i \in E(I_l), e_i \in E(I_k),$$

тобто для структурного елемента його текстуровані дескриптори на кадрах відеопослідовності, на яких він присутній, приблизно збігаються.

Окремим об'єктам на зображенні I відповідають підмножини структурних елементів множини $E(I)$, що випливає з визначення структурного елемента і властивості локалізації. Таким чином, кожен об'єкт на зображенні описується множиною структурних елементів:

$$O_k = \{e_j : \forall j\, (x_j, y_j) \in \Omega_k^I\}_j, \tag{2}$$

де Ω_k – область зображення, яку займає *k*-тий об'єкт. Відповідно для зображення I_l, опис об'єкта, що належить відеопослідовності, буде виглядати наступним чином:

$$O_k^l = \{e_j : \forall j\, (x_j, y_j)^l \in \Omega_k^{I_l}\}_j,$$

а опис об'єкта на всій відеопослідовності описується виразом:

$$O_k = \{e_j : \forall j\, (x_j, y_j)^l \in \Omega_k^{I_l}\}_{j,l},$$

що з точністю до наявності індексу кадра збігається з виразом для опису об'єкта на окремому кадрі. Зокрема, це означає, що якщо у відеопослідовності *V* визначена множина $E(V)$ та для кадру *l* і об'єкту *k* визначено множину O_k^l і оскільки кожен структурний елемент визначено на послідовності кадрів відеопослідовності, то справедливе відношення:

$$O_k^l \cup O_k^{l+1} \neq \varnothing. \tag{3}$$

Тобто, якщо на окремому кадрі відеопослідовності для об'єкта визначено його опис у вигляді множини структурних елементів, то визначено частину опису даного об'єкта на наступному кадрі. Невідповідність у загальному випадку описів O_k^l и O_k^{l+1} пов'язана з тим, що при русі об'єкта окремі його області можуть зникати з поля зору камери, а інші, навпаки, з'являтися. Дана зміна зовнішнього виду об'єкта призводить до зникнення одних структурних елементів і появи інших. Таким чином, від кадра до кадра множина структурних елементів, що відповідають одному об'єкту модифікується, але завжди існують структурні елементи, присутні на двох сусідніх кадрах відеопослідовності.

Застосування опису об'єктів у вигляді множини структурних елементів, визначених на всій відеопослідовності, що розглядається, дозволяє значно спростити процедуру відстеження (трекінгу) об'єктів, у тому числі і в разі їх колізій (перекриття). Визначення модифікації опису об'єкта від кадра до кадра дозволяє відстежувати зміни його форми, розміру і текстури на всій відеопослідовності.

При русі об'єктів взаємне розташування сусідніх структурних елементів (тобто структура), що відносяться до одного об'єкта, як правило, змінюється не значно, оскільки зовнішній вигляд об'єктів на сусідніх кадрах змінюється не суттєво. При застосуванні відношення 1 до сусідніх структурних елементів одного об'єкта повинна дотримуватися умова:

$$\vec{s}_i^{\,l} \approx \vec{s}_j^{\,l} : e_i, e_j \in O_k, (e_i, e_j) \in \mathrm{N}_l, \tag{4}$$

де O_k – опис *k*-того об'єкта, e_i, e_j – структурні елементи, що належать опису *k*-того об'єкта і є сусідами, N_l – множина всіх пар структурних елементів, які є сусідніми один одному на *i*-ом кадрі. Таким чином, окремим об'єктам в кадрі відповідають підмножини структурних елементів, для кожного з яких виконується умова збігу зміщення між кадрами для сусідів

$$\vec{s}_i^{\,l} \approx \vec{s}_j^{\,l} : \forall j : (e_i, e_j) \in \mathrm{N}_l \,. \tag{5}$$

Зрозуміло, якщо два об'єкти в кадрі знаходяться поруч і однаково зміщуються, то вони можуть відслідковуватися як один об'єкт до тих пір, поки вони не розійдуться. Однак при зближенні різних об'єктів їх злиття не повинно відбуватися внаслідок відношення 3. Властивості умови 5 надалі будуть використані для виділення (детектування) і відстеження (трекінгу) рухомих об'єктів на кадрах відеопослідовності.

Як видно з вищевикладеного, використання представлення зображень і відеопослідовностей у вигляді множини структурних елементів має переваги для вирішення задач детектування і відстеження рухомих об'єктів. Для його використання необхідно для кожного кадра провести процедуру виділення структурних елементів і визначення зміщення між кадрами для них.

Детектування і трекінг об'єктів, що рухаються

Детектування рухомих об'єктів здійснюється за рахунок виділення підмножин структурних елементів на окремому кадрі, які задовольняють умові (4). Позначимо $S^l = \{S_i^l\}_i$ множину задетектованих об'єктів на *l*-ому кадрі. Кожен виділений об'єкт являє собою підмножину всіх структурних об'єктів у поточному кадрі. Процедура виділення об'єктів, що рухаються, проводиться на кожному кадрі. Дана процедура повинна мати наступну властивість – результат виділення об'єктів повинен бути незалежним від результатів, отриманих на попередніх кадрах. Дана властивість дозволяє розглядати детектування і трекінг об'єктів як єдиний процес і здійснювати відстеження об'єктів з урахуванням результатів детектування.

У разі використання запропонованого подання для опису відеопослідовності кожен структурний елемент заданий на послідовності кадрів. Так само як і у випадку з одиничним зображенням, окремим об'єктам відповідають підмножини структурних елементів. Таким чином, завдання відстеження цільових об'єктів зводиться до пошуку відповідних їм підмножин структурних елементів.

Оскільки при обробці відеопослідовності аналіз зображень здійснюється послідовно (кадр за кадром, у міру надходження зображень з відеопотоку) – трекінг об'єктів полягає у модифікації їх опису при переході від кадра до кадра

$$O_k^l \rightarrow O_k^{l+1} .$$

Модифікація множини елементів, що описує об'єкт, є функцією перетину і об'єднання опису об'єкта на попередньому кадрі з множинами виділених об'єктів на поточному кадрі

$$O_k^{l+1} = \Phi(O_k^l \cup \bigcup_i S_i^{l+1}, O_k^l \cap \bigcup_i S_i^{l+1}) \tag{6}$$

Використання виділених на поточному кадрі об'єктів дозволяє відстежувати зміни опису об'єкта від кадра до кадра і об'єднати детектування і трекінг об'єктів у єдиний процес. Функція (6) повинна мати наступні властивості.

Властивість 1. Якщо виділений об'єкт не перетинається ні з одним об'єктом, що відслідковується – породжується новий об'єкт, що відстежується,

$$\exists S_k^{l+1} : S_k^{l+1} \cap \bigcup_i O_i^{l+1} = \varnothing \Rightarrow O_n : O_n^{l+1} = S_k^{l+1},$$

де O_n – новий об'єкт, що відстежується.

Властивість 2. Якщо об'єкт, що відстежується, не перетинається ні з одним виділеним об'єктом – даний об'єкт, що відстежується, видаляється

$$\exists O_k^{l+1} : O_k^{l+1} \cap \bigcup_i S_i^{l+1} = \varnothing \Rightarrow \mathit{delete\,objekt}\ k,$$

де O_k^l – опис *k*-ого об'єкта, що відстежується,на *l*-ом кадрі; S_k^l – опис *k*-ого виділеного на *l*-ому кадрі об'єкта.

Властивість 3. При модифікації $O_k^l \to O_k^{l+1}$ повинні зберігатися властивості опису об'єктів (3) і (4).

У даній роботі не розглядаються конкретні реалізації функції (6) і процедури виділення множини задетектованих об'єктів на *l*-му кадрі $S^l = \{S_i^l\}_i$, оскільки вони залежать від супутніх факторів – способу екстрагування структурних елементів, вхідних даних, поставлених задач і т.д.

Розпізнавання виділених об'єктів

Опис об'єктів на окремому кадрі у вигляді множини структурних елементів (2) дозволяє здійснювати їх розпізнавання. Дійсно, оскільки кожен структурний елемент характеризується 1) своїм дескриптором, який описує його яскравості і текстури властивості, і 2) координатами, які визначають структуру об'єкта, то інформації, що міститься в описі (2) досить для розпізнавання. Для розпізнавання k-ого об'єкта на l-му кадрі необхідно сформувати його дескриптор $D(O_k^l)$, який є функцією від опису (2). Наприклад, опис (2) дозволяє використовувати як дескриптор граф, кожній вершині якого відповідає структурний елемент, а ребра описують відносини між ними [35]. Однак, можливі й інші варіанти дескрипторів об'єктів.

Типи структурних елементів

Структурні елементи, які використовуються для опису зображень відеопослідовності, можна розділити на два типи за способом їх екстракції на поточному кадрі:

1. Структурні елементи виділяються на основі використання статичного зображення (використовується метод для пошуку областей/точок, які мають задані властивості).
2. Структурні елементи виділяються на основі пошуку областей/точок відповідних структурним елементам на попередньому кадрі.

Властивості структурних елементів даних типів відображені в таблиці 1.

Таблиця 1

Тип структурного елемента	Текстурні властивості	Спосіб визначення міжкадрового зміщення
1	Структурні елементи мають певні текстурні властивості, що залежать від способу їх екстракції.	Для пошуку відповідності між структурними елементами на сусідніх кадрах необхідно використовувати спеціальну процедуру.
2	Як структурний елемент може виступати будь-яка область зображення.	Відповідність між структурними елементами сусідніх кадрів формується автоматично в процесі їх екстракції.

У разі, якщо завдання інтелектуального аналізу відеопослідовності обмежуються детектуванням і трекінгом рухомих об'єктів, то доцільно використовувати структурні елементи другого типу, оскільки вони вимагають менше витрат для їх екстракції і пошуку міжкадрового зміщення.

Для вирішення завдань розпізнавання більше підходять структурні елементи першого типу, оскільки зменшується свавілля в їх екстракції і, відповідно, дескриптори об'єктів менш схильні до змін, що залежать від їх положення і виділених структурних елементів.

Слід зауважити, що запропонований формалізм дозволяє спільне використання структурних елементів, які вимагають різних способів екстракції і належать різним типам. Дана властивість дозволяє 1) зменшити витрати для визначення межкадрового зміщення для структурних елементів першого типу за рахунок даних, отриманих при екстракції структурних елементів другого типу; 2) зменшити свавілля при екстракції структурних елементів другого типу за рахунок даних, отриманих при екстракції структурних елементів першого типу. Запропонований формалізм не конкретизує вид дескриптора для структурних елементів, що дозволяє використовувати різні дескриптори для структурних елементів різної природи (за способом екстракції).

Висновки

У даній роботі вперше зроблено формальний опис зображень і видеопослідовностей у вигляді множини структурних елементів, дано визначення опису виділених об'єктів і об'єктів, що відслідковуються, розглянуто властивості структурних елементів, що належать одному об'єкту, властивості опису виділених об'єктів, властивості описів об'єктів, що відслідковуються, властивості функції перетворення/модифікації опису об'єктів, що відслідковуються, від кадра до кадра, яка необхідна для здійснення трекінгу.

У результаті проведеного дослідження зроблено такі висновки:

1. Використання опису зображення у вигляді множини структурних елементів дозволяє формувати опис окремого об'єкта не тільки на окремому кадрі, а й на послідовності кадрів. Дана властивість дозволяє аналізувати динамічні властивості об'єктів.

2. Запропонований структурний опис зображень і об'єктів дозволяє проводити виділення рухомих об'єктів і їх відстеження (трекінг) на основі використання єдиного процесу.

3. Розгляд зображення як множини структурних елементів дозволяє формувати описи окремих об'єктів, які дозволяють проводити їх розпізнавання.

4. Запропонований спосіб опису зображень і об'єктів дозволяє як структурні елементи використовувати ділянки зображень з різними властивостями, що мають дескриптори різної природи.

Подальші роботи будуть присвячені розвитку теоретичних положень щодо використання описів зображень і видеопослідовностей у вигляді множини структурних елементів для інтелектуально аналізу відеоданих і їх практичного застосування.

Література

1. Jain R. On the analysis of accumulative difference pictures from image sequences of real world scenes / Jain R., Nagel H. // IEEE Trans. Patt. Analy. Mach. Intell. 1979. - 1, 2. - P. 206–214.
2. Torre Frade de la F. Moving object detection and tracking system: a real-time implementation / F. de la Torre Frade, E.M. Marroquín, E.S. Pérez, J.A.M. Moreno // Seizième colloque gretsi — 15-19 septembre, 1997. — P. 375-378.
3. Yilmaz A. Object Tracking: A Survey / Alper Yilmaz, Omar Javed, Mubarak Shah // ACM Computing Surveys. – 2006. – Vol. 38. – No. 4. – Article 13. – P.1-45.
4. Papadourakis V. Multiple objects tracking in the presence of long-term occlusions / V. Papadourakis, A. Argyros // Computer Vision and Image Understanding. — 2010 — vol.114. – P. 835–846.
5. Wang Y. Moving Object Tracking in Video / Y. Wang, J.F. Doherty, R.E. Van Dyck // Applied Imagery Pattern Recognition Workshop/ - 2000.
6. Sechidis L.A. Low-Level Tracking Of Multiple Objects / L.A. Sechidis, P. Patias, V. Tsioukas // IAPRS. - 2000. - P.237-240.
7. Wren C. Pfinder: Real-time tracking of the human body / C. Wren, A. Zarbayejani, P. Entland A. // IEEE Trans. Patt. Analy. Mach. Intell. 19, 7.- 1997.- P.780–785.
8. Arroyo R. Expert Video-Surveillance System for Real-Time Detection of Suspicious Behaviors in Shopping Malls / R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza, J. Almazán // Expert Systems with Applications. - Vol. 42. - Issue 21. - November 2015. - P. 7991-8005.
9. Dhananjaya B. Detection of Objects in Aerial Videos for Object Extraction and Tracking for UAV Applications / B. Dhananjaya, B. Rama Murthy, P. Thimmaiah // International Journal of Computer Applications (0975 – 8887). - Vol. 112 – No 12. - February 201. - P.37-42.
10. Ramirez A.A., Chouikha M. A New Algorithm for Tracking Objects in Videos of Cluttered Scenes / A.A. Ramirez, M. Chouikha // International Journal of Information Technology, Modeling and Computing (IJITMC) – May, 2013. - vol.1, No.2, – P.72-83.
11. Collins R.T. On-Line Selection of Discriminative Tracking Features / R.T. Collins, Y.Liu // Pattern Analysis and Machine Intelligence. – 2005. – vol. 27, issue 10. – P. 1631 – 1643.
12. Bolme D.S. Visual Object Tracking using Adaptive Correlation Filters / D.S. Bolme, J.R. Beveridge, B.A. Draper, Y.M. Lui // Computer Vision and Pattern Recognition. — June, 2010. — P. 2544 – 2550.
13. Xu M. A Two-Level Approach for Visual Tracking / M. Xu, J. Yang, Y. Pen // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. - Vol. 1. - 2005. - P. 688-695.
14. Godec M. Hough-based Tracking of Non-rigid Objects / M. Godec, P.M. Roth, H. Bischof // In Proc. International Conference on Computer Vision (ICCV). — 2011. – P. 81 – 88.
15. Schunck B.G. Image Flow Segmentation and Estimation by Constraint Line Clustering / B.G. Schunck // IEEE Transactions on Pattern Analysis and Machine Intelligence. - Vol.11, Issue 10. - 2002. - P. 1010 - 1027.
16. Brox T. Object Segmentation by Long Term Analysis of Point Trajectories / T. Brox, J. Malik // 11th European conference on Computer vision. - 2010. - P. 282-295.
17. Vojir T. The Enhanced Flock of Trackers / T. Vojir, J. Matas // Registration and Recognition in Images and Videos. - January, 2014. - P. -113-136.
18. Matas J. Robustifying the Flock of Trackers / J. Matas, T. Vojir // Computer Vision Winter Workshop. - 2001. - P. 91-97.
19. Kölsch M. Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration / M. Kölsch, M. Turk // Conference on Computer Vision and Pattern Recognition Workshop. -Vol. 10. - 2004. - P.158.
20. Maresca M.E. Clustering Local Motion Estimates for Robust and Efficient Object Tracking / M.E. Maresca, A. Petrosino // ECCV Workshops. - 2014. - P.244-253.
21. Kalal Z. Online learning of robust object detectors during unstable tracking / Z. Kalal, J. Matas, K. Mikolajczyk // 3rd On-line Learning for Computer Vision Workshop. - Kyoto, Japan. - 2009. - P. 1417-1424.

22. Kalal Z. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints / Z. Kalal, J. Matas, K. Mikolajczyk // Published at the 23rd IEEE Conference on Computer Vision and Pattern Recognition. - San Francisco, CVPR. - June 13-18. - 2010. - P.49 - 56.
23. Kalal Z. Face-Tld: Tracking-Learning-Detection Applied To Faces / Z. Kalal, J. Matas, K. Mikolajczyk // 17th IEEE International Conference on Image Processing (ICIP). - 2010. - P. 3789-3792.
24. Kalal Z. Forward-Backward Error: Automatic Detection of Tracking Failures / Z. Kalal, J. Matas, K. Mikolajczyk // International Conference on Pattern Recognition. - Istambul, Turkey. - 2010. - P. 2756-2759.
25. Kalal Z. Tracking-Learning-Detection / Z. Kalal, J. Matas, K. Mikolajczyk // IEEE Transactions On Pattern Analysis And Machine Intelligence. - vol. 6, №. 1. - january, 2010. - P.1409-1422.
26. Babenko B. Visual Tracking with Online Multiple Instance Learning / B. Babenko, M.-H. Yang, S. Belongie // IEEE Conference on Computer Vision and Pattern Recognition. - 2009.
27. Babenko B. Robust Object Tracking with Online Multiple Instance Learning / B. Babenko, M.-H. Yang, S. Belongie // IEEE Transactions on Pattern Analysis & Machine Intelligence. - Vol. 33, No. 08. - August, 2011. - P.1619-1632.
28. Ross D.A. Incremental Learning for Robust Visual Tracking / D.A. Ross, J. Lim, R.-S. Lin, M.-H. Yang // International Journal of Computer Vision. - Vol. 77, Issue 1. - 2008. - P.125–141.
29. Lim J. Incremental Learning for Visual Tracking / J. Lim, D. Ross, R.-S. Lin, M.-H. Yang // Advances in Neural Information Processing Systems. - 2004. - P. 793-800.
30. Piao S. Multi-Object Tracking Based on Tracking-Learning-Detection Framework / S. Piao, K. Berns // Field and Assistive Robotics─Advances in Systems and Algorithms. ─ 2014. ─ P. 74-87.
31. Sunil T.N. Multiple Moving Object Detection and Tracking using Harr Features with Smart Video Surveillance System / Sunil T.N., Ravikumar A.V. // International Journal of Engineering Research & Technology. - Vol. 3, Issue 6. - 2014.
32. Агарков А.В. Трекинг объектов на основе использования структурного описания изображений / А.В. Агарков // Системы и средства искусственного интеллекта. ─ 2014. ─ № 1. ─ С. 133-136.
33. Агарков А.В. Выделение и отслеживание объектов на основе использования анализа движения / А.В. Агарков // Искусственный интеллект. ─ 2015. ─ № 1-2. ─ С. 28-35.
34. Агарков А.В. Использование структурного описания изображений для детектирования и трекинга движущихся объектов / А.В. Агарков // Искусственный интеллект. Интеллектуальные транспортные системы: материалы Международной научно-технической конференции. – Брест: УО БрГТУ, 2016 – С. 26–28.
35. Агарков А.В. Поиск объектов на изображениях с помощью графов / А.В. Агарков // Искусственный интеллект. ─ 2012. ─ № 4. ─ С. 105-119.

Literatura

1. Jain R. On the analysis of accumulative difference pictures from image sequences of real world scenes / Jain R., Nagel H. // IEEE Trans. Patt. Analy. Mach. Intell. 1979. - 1, 2. - P. 206–214.
2. Torre Frade de la F. Moving object detection and tracking system: a real-time implementation / F. de la Torre Frade, E.M. Marroquín, E.S. Pérez, J.A.M. Moreno // Seizième colloque gretsi — 15-19 septembre, 1997. — P. 375-378.
3. Yilmaz A. Object Tracking: A Survey / Alper Yilmaz, Omar Javed, Mubarak Shah // ACM Computing Surveys. – 2006. – Vol. 38. – No. 4. – Article 13. – P.1-45.
4. Papadourakis V. Multiple objects tracking in the presence of long-term occlusions / V. Papadourakis, A. Argyros // Computer Vision and Image Understanding. — 2010 — vol.114. – P. 835–846.
5. Wang Y. Moving Object Tracking in Video / Y. Wang, J.F. Doherty, R.E. Van Dyck // Applied Imagery Pattern Recognition Workshop/ - 2000.
6. Sechidis L.A. Low-Level Tracking Of Multiple Objects / L.A. Sechidis, P. Patias, V. Tsioukas // IAPRS. - 2000. - P.237-240.
7. Wren C. Pfinder: Real-time tracking of the human body / C. Wren, A. Zarbayejani, P. Entland A. // IEEE Trans. Patt. Analy. Mach. Intell. 19, 7.- 1997.- P.780–785.
8. Arroyo R. Expert Video-Surveillance System for Real-Time Detection of Suspicious Behaviors in Shopping Malls / R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza, J. Almazán // Expert Systems with Applications. - Vol. 42. - Issue 21. - November 2015. - P. 7991-8005.
9. Dhananjaya B. Detection of Objects in Aerial Videos for Object Extraction and Tracking for UAV Applications / B. Dhananjaya, B. Rama Murthy, P. Thimmaiah // International Journal of Computer Applications (0975 – 8887). - Vol. 112 – No 12. - February 201. - P.37-42.

10. Ramirez A.A., Chouikha M. A New Algorithm for Tracking Objects in Videos of Cluttered Scenes / A.A. Ramirez, M. Chouikha // International Journal of Information Technology, Modeling and Computing (IJITMC) – May, 2013. - vol.1, No.2, – P.72-83.
11. Collins R.T. On-Line Selection of Discriminative Tracking Features / R.T. Collins, Y.Liu // Pattern Analysis and Machine Intelligence. – 2005. – vol. 27, issue 10. – P. 1631 – 1643.
12. Bolme D.S. Visual Object Tracking using Adaptive Correlation Filters / D.S. Bolme, J.R. Beveridge, B.A. Draper, Y.M. Lui // Computer Vision and Pattern Recognition. — June, 2010. — P. 2544 – 2550.
13. Xu M. A Two-Level Approach for Visual Tracking / M. Xu, J. Yang, Y. Pen // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. - Vol. 1. - 2005. - P. 688-695.
14. Godec M. Hough-based Tracking of Non-rigid Objects / M. Godec, P.M. Roth, H. Bischof // In Proc. International Conference on Computer Vision (ICCV). — 2011. – P. 81–88.
15. Schunck B.G. Image Flow Segmentation and Estimation by Constraint Line Clustering / B.G. Schunck // IEEE Transactions on Pattern Analysis and Machine Intelligence. - Vol.11, Issue 10. - 2002. - P. 1010 - 1027.
16. Brox T. Object Segmentation by Long Term Analysis of Point Trajectories / T. Brox, J. Malik // 11th European conference on Computer vision. - 2010. - P. 282-295.
17. Vojir T. The Enhanced Flock of Trackers / T. Vojir, J. Matas // Registration and Recognition in Images and Videos. - January, 2014. - P. -113-136.
18. Matas J. Robustifying the Flock of Trackers / J. Matas, T. Vojir // Computer Vision Winter Workshop. - 2001. - P. 91-97.
19. Kölsch M. Fast 2D Hand Tracking with Flocks of Features and Multi-Cue Integration / M. Kölsch, M. Turk // Conference on Computer Vision and Pattern Recognition Workshop. -Vol. 10. - 2004. - P.158.
20. Maresca M.E. Clustering Local Motion Estimates for Robust and Efficient Object Tracking / M.E. Maresca, A. Petrosino // ECCV Workshops. - 2014. - P.244-253.
21. Kalal Z. Online learning of robust object detectors during unstable tracking / Z. Kalal, J. Matas, K. Mikolajczyk // 3rd On-line Learning for Computer Vision Workshop. - Kyoto, Japan. - 2009. - P. 1417-1424.
22. Kalal Z. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints / Z. Kalal, J. Matas, K. Mikolajczyk // Published at the 23rd IEEE Conference on Computer Vision and Pattern Recognition. - San Francisco, CVPR. - June 13-18. - 2010. - P. 49-56.
23. Kalal Z. Face-Tld: Tracking-Learning-Detection Applied To Faces / Z. Kalal, J. Matas, K. Mikolajczyk // 17th IEEE International Conference on Image Processing (ICIP). - 2010. - P. 3789-3792.
24. Kalal Z. Forward-Backward Error: Automatic Detection of Tracking Failures / Z. Kalal, J. Matas, K. Mikolajczyk // International Conference on Pattern Recognition. - Istambul, Turkey. - 2010. - P. 2756-2759.
25. Kalal Z. Tracking-Learning-Detection / Z. Kalal, J. Matas, K. Mikolajczyk // IEEE Transactions On Pattern Analysis And Machine Intelligence. - vol. 6, №. 1. - january, 2010. - P.1409-1422.
26. Babenko B. Visual Tracking with Online Multiple Instance Learning / B. Babenko, M.-H. Yang, S. Belongie // IEEE Conference on Computer Vision and Pattern Recognition. - 2009.
27. Babenko B. Robust Object Tracking with Online Multiple Instance Learning / B. Babenko, M.-H. Yang, S. Belongie // IEEE Transactions on Pattern Analysis & Machine Intelligence. - Vol. 33, No. 08. - August, 2011. - P.1619-1632.
28. Ross D.A. Incremental Learning for Robust Visual Tracking / D.A. Ross, J. Lim, R.-S. Lin, M.-H. Yang // International Journal of Computer Vision. - Vol. 77, Issue 1. - 2008. - P.125–141.
29. Lim J. Incremental Learning for Visual Tracking / J. Lim, D. Ross, R.-S. Lin, M.-H. Yang // Advances in Neural Information Processing Systems. - 2004. - P. 793-800.
30. Piao S. Multi-Object Tracking Based on Tracking-Learning-Detection Framework / S. Piao, K. Berns // Field and Assistive Robotics─Advances in Systems and Algorithms. ─ 2014. ─ P. 74-87.
31. Sunil T.N. Multiple Moving Object Detection and Tracking using Harr Features with Smart Video Surveillance System / Sunil T.N., Ravikumar A.V. // International Journal of Engineering Research & Technology. - Vol. 3, Issue 6. - 2014.
32. Agarkov A.V. Treking ob"yektov na osnove ispol'zovaniya strukturnogo opisaniya izobrazheniy / A. V. Agarkov // Sistemy i sredstva iskusstvennogo intellekta. - 2014. - № 1. - S. 133-136.
33. Agarkov A.V. Vydeleniye i otslezhivaniye ob"yektov na osnove ispol'zovaniya analiza dvizheniya / A. V. Agarkov // Iskusstvennyy intellekt. - 2015. - № 1-2. - S. 28-35.
34. Agarkov A.V. Ispol'zovaniye strukturnogo opisaniya izobrazheniy dlya detektirovaniya i trekinga dvizhushchikhsya ob"yektov / A. V. Agarkov // Iskusstvennyy intellekt. Intellektual'nyye transportnyye sistemy: materialy Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii. – Brest: UO BrGTU, 2016 – S. 26 – 28.
35. Agarkov A.V. Poisk ob"yektov na izobrazheniyakh s pomoshch'yu grafov / A.V. Agarkov // Iskusstvennyy intellekt. - 2012. - № 4. - S. 105-119.

RESUME

A.V. Agarkov

Application of images structural description for solving problems of intellectual analysis of video sequences

In this paper, we consider the use of the description of images and video sequences in the form of a set of structural elements for solving problems of detection, tracking and recognition of moving objects. Theoretical studies of this problem have been carried out. A formal description of images and video sequences in the form of a number of structural elements is given, a definition of the description of detected and tracked objects is given, properties of structural elements belonging to one object, properties of detected objects description, properties of tracked objects descriptions, properties of the transformation/modification function of tracked objects description from frame to frame, which is necessary for tracking, are considered.

To describe the images, instead of the two-dimensional limited function of two variables, it is suggested to use a set of structural elements. Each structural element occupies a certain area of the image, is characterized by the coordinates of the center and the descriptor, which describes the brightness and texture properties of the area. Objects in the image correspond to subsets of this set of structural elements.

When proceeding to consideration of video sequences, a single structural element is determined on a sequence of frames. Those, the parameters characterizing the single structural element depend on the frame. Since it is assumed that a particular structural element corresponds to a specific object (or part of it), its texture properties should not significantly differ for the entire sequence of frames.

Objects, as in the case of a single image, are described as sets of structural elements. The difference is that each structural element is specified on a sequence of frames, which leads to the description of the object for the entire video sequence (or its part).

Since the structure of objects for two neighboring frames does not change significantly, the interframe shift (change of coordinates from frame to frame) of their structural objects must coincide (at least for neighbors). This property is used to detect moving objects on the current frame.

Since each structural element is specified in a sequence of frames, the intersection of sets, which is a descriptions of one object on two consecutive frames, is not zero. This property allow to consider the tracking process as a modification of the set of structural elements that describe the object, from frame to frame. This process is formally described as a function, depending on the descriptions of objects on the previous frame and objects detected on the current frame. Using objects selected on the current frame in the tracking.

Since each structural element is specified in a sequence of frames, the suppression of their sets, which is a description of one object on two consecutive frames, is not zero. This property allows us to consider the tracking process as a modification of the set of structural elements that describe the object, from frame to frame. This process is formally described as a function, depending on the descriptions of objects on the previous frame and those selected on the current frame of objects. Using objects detected on the current frame in the tracking process allows to track changes in the description of the object from frame to frame and combine the detection and tracking of objects into a single process.

The description of objects on a separate frame in the form of a set of structural elements allows them to be recognized. This follows from the fact that the information contained in this set of structural elements is sufficient to describe the texture and structure of the object.

The structural elements that can be used to form the proposed description are divided into two types, based on the method of their extraction on the current image of the video sequence. The properties of structural elements of these types and the scope of the descriptions based on them are considered.

The proposed formalism allows the use of structural elements of different types in one description, which makes it possible to compensate for some of the shortcomings inherent in each type. This formalism does not specify the type of descriptor for structural elements, which allows using different descriptors for structural elements of different nature (by extraction method).

Надійшла до редакції 04.05.2017

